UNITED STATES PATENT AND TRADEMARK OFFICE

UNITED STATES DEPARTMENT OF COMMERCE
United States Patent and Trademark Office
Address: COMMISSIONER FOR PATENTS
P.O. Box 1450
Alexandria, Virginia 22313-1450
www.uspto.gov

APPLICATION NO.	FILING DATE	FIRST NAMED INVENTOR	ATTORNEY DOCKET NO.	CONFIRMATION NO.
10/723,049	11/26/2003	Xian-Sheng Hua	MS1-1744US	7823

22801 7590 05/12/2009
LEE & HAYES, PLLC
601 W. RIVERSIDE AVENUE
SUITE 1400
SPOKANE, WA 99201

EXAMINER	
GISHNOCK, NIKOLAI A	

ART UNIT	PAPER NUMBER
3715	

MAIL DATE	DELIVERY MODE
05/12/2009	PAPER

Please find below and/or attached an Office communication concerning this application or proceeding.

The time period for reply, if any, is set in the attached communication.

Office Action Summary	**Application No.** 10/723,049	**Applicant(s)** HUA ET AL.	
	Examiner NIKOLAI A. GISHNOCK	**Art Unit** 3715	

-- The MAILING DATE of this communication appears on the cover sheet with the correspondence address --

Period for Reply

A SHORTENED STATUTORY PERIOD FOR REPLY IS SET TO EXPIRE _3_ MONTH(S) OR THIRTY (30) DAYS, WHICHEVER IS LONGER, FROM THE MAILING DATE OF THIS COMMUNICATION.

- Extensions of time may be available under the provisions of 37 CFR 1.136(a). In no event, however, may a reply be timely filed after SIX (6) MONTHS from the mailing date of this communication.
- If NO period for reply is specified above, the maximum statutory period will apply and will expire SIX (6) MONTHS from the mailing date of this communication.
- Failure to reply within the set or extended period for reply will, by statute, cause the application to become ABANDONED (35 U.S.C. § 133). Any reply received by the Office later than three months after the mailing date of this communication, even if timely filed, may reduce any earned patent term adjustment. See 37 CFR 1.704(b).

Status

1)☒ Responsive to communication(s) filed on _05 February 2009_.

2a)☒ This action is **FINAL**. 2b)☐ This action is non-final.

3)☐ Since this application is in condition for allowance except for formal matters, prosecution as to the merits is closed in accordance with the practice under *Ex parte Quayle*, 1935 C.D. 11, 453 O.G. 213.

Disposition of Claims

4)☒ Claim(s) _1,5-8,10-15,17-19 and 21-45_ is/are pending in the application.

4a) Of the above claim(s) _____ is/are withdrawn from consideration.

5)☐ Claim(s) _____ is/are allowed.

6)☒ Claim(s) _1,5-8,10-15,17-19 and 21-45_ is/are rejected.

7)☐ Claim(s) _____ is/are objected to.

8)☐ Claim(s) _____ are subject to restriction and/or election requirement.

Application Papers

9)☐ The specification is objected to by the Examiner.

10)☒ The drawing(s) filed on _26 November 2003_ is/are: a)☒ accepted or b)☐ objected to by the Examiner.

Applicant may not request that any objection to the drawing(s) be held in abeyance. See 37 CFR 1.85(a).

Replacement drawing sheet(s) including the correction is required if the drawing(s) is objected to. See 37 CFR 1.121(d).

11)☐ The oath or declaration is objected to by the Examiner. Note the attached Office Action or form PTO-152.

Priority under 35 U.S.C. § 119

12)☐ Acknowledgment is made of a claim for foreign priority under 35 U.S.C. § 119(a)-(d) or (f).

a)☐ All b)☐ Some * c)☐ None of:

1.☐ Certified copies of the priority documents have been received.

2.☐ Certified copies of the priority documents have been received in Application No. _____.

3.☐ Copies of the certified copies of the priority documents have been received in this National Stage application from the International Bureau (PCT Rule 17.2(a)).

* See the attached detailed Office action for a list of the certified copies not received.

Attachment(s)

1) ☐ Notice of References Cited (PTO-892)

2) ☐ Notice of Draftsperson's Patent Drawing Review (PTO-948)

3) ☐ Information Disclosure Statement(s) (PTO/SB/08) Paper No(s)/Mail Date _____.

4) ☐ Interview Summary (PTO-413) Paper No(s)/Mail Date. _____ .

5) ☐ Notice of Informal Patent Application

6) ☐ Other: _____.

DETAILED ACTION

In response to Applicant's remarks filed 2/5/2009, claims 2-4, 9, 16, & 20 are cancelled. Claims 1, 5, 6-8, 10-15, 17-19, & 21-45 are pending.

Claim Rejections - 35 USC § 103

1. The following is a quotation of 35 U.S.C. 103(a) which forms the basis for all obviousness rejections set forth in this Office action:

> (a) A patent may not be obtained though the invention is not identically disclosed or described as set forth in section 102 of this title, if the differences between the subject matter sought to be patented and the prior art are such that the subject matter as a whole would have been obvious at the time the invention was made to a person having ordinary skill in the art to which said subject matter pertains. Patentability shall not be negatived by the manner in which the invention was made.

2. The factual inquiries set forth in *Graham* **v.** *John Deere Co.*, 383 U.S. 1, 148 USPQ 459 (1966), that are applied for establishing a background for determining obviousness under 35 U.S.C. 103(a) are summarized as follows:

1. Determining the scope and contents of the prior art.
2. Ascertaining the differences between the prior art and the claims at issue.
3. Resolving the level of ordinary skill in the pertinent art.
4. Considering objective evidence present in the application indicating obviousness or nonobviousness.

3. Claims 1, 5-8, 10, 17, 18, 23-25, 28, 29, 32, 33, 40, & 41 are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky (US 5,782,692), hereinafter known as Stelovsky, in view of Wang, (US 2002/0133764 A1), hereinafter known as Wang, further in view of Hansen et al. (US 2002/0038456 A1), hereinafter known as Hansen, Umeda (US 5,453,570 A), hereinafter known as Umeda, Golin (US 5,990,980 A), hereinafter known as Golin, and Osberger (US 6,670,963), hereinafter known as Osberger.

4. Stelovsky teaches a processor-readable medium comprising executable instructions for personalizing karaoke (Column 1, Lines 54-67), comprising: segmenting visual content to produce a plurality of sub-shots, where the instructions for segmenting visual content segment video, and segmenting music to produce a plurality of music sub-clips (multimedia presentation track consisting of video, audio, and text display is segmented with respect to specific beginning and ending points, Column 3, Lines 27-65); selecting important sub-shots from within the plurality of sub-shots (Column 3, Lines 52-60; it is understood that the selected sub-shots are important to the user); and displaying at least some of the plurality of sub-shots as a background to lyrics associated with the plurality of music sub-clips ("Karaoke Game" presentation has synchronized video and instrumental sound tracks, Column 9, Lines 15-21; the text can be superimposed on the video, Column 10, Lines 5-6). [Claim 1].

5. Stelovsky teaches a processor-readable medium comprising instructions for providing lyrics for integrating lyrics, music, and video content suitable for karaoke, comprising instructions for: receiving a request for a file associated with a specific song (clicking on a word in the text track, Column 14, Lines 42-48), wherein the file comprises music, lyrics, and timing values (The time-dependent sequence is composed of tracks that are synchronized with respect to a common time axis {hereinafter "multimedia presentation"}. The basic track consists of video display images and is synchronized with at least one other track that consists of audio or text display, 3:31-35; The multimedia presentation is segmented with respect to specific beginning and ending points of segments on the time axis, i.e. there are one or more points of time that partition the time axis into time segments, 3:52-55), and fulfilling the request by sending the file associated with the specified song (connection is established with a remote on-line service, search query initiated, and results are displayed, Column 14, Lines 42-48), segmenting visual content to produce a plurality of sub-shots of a length corresponding to the music sub-clips

(multimedia presentation track consisting of video, audio, and text display is segmented with respect to specific beginning and ending points, Column 3, Lines 27-65), and outputting the plurality of music sub-clips together with corresponding sub-shots of visual content, which is configured as a background to the lyrics associated with the music sub-clips ("Karaoke Game" presentation has synchronized video and instrumental sound tracks, Column 9, Lines 15-21; the text can be superimposed on the video, Column 10, Lines 5-6) [Claim 23].

6. Stelovsky teaches a personalized karaoke device, comprising: a music analyzer configured to create music sub-clips of varying lengths according to a song (Segmentation Authoring System {SAS} facilitates the identification of points in time where a segment starts and ends, Column 5, Line 62 to Column 6, Line 2; multimedia presentation track consisting of video, audio, and text display is segmented with respect to specific beginning and ending points, Column 3, Lines 27-65); a visual content analyzer configured to define and select visual content sub-shots (Using SAS, the author partitions the multimedia presentation into time segments according to predominant time units, e.g., measures of song, sound bites, or action sequences in a movie, Column 6, Lines 51-54); a lyric formatter configured to time delivery of syllables of lyrics of the song (evaluation feedback of user's input includes visualization of differences in pronunciation patterns, processes involved in generating {human} speech, such as positions of the tongue and air flow patterns, Column 14, Lines 52-59; it is inherent that the speech analysis as disclosed could recognize syllables and sentences, which are pronunciation patterns); sections of the text track are linked to the time segments, Column 6, Line 55); and a composer configured to assemble the music sub-clips with the visual content sub-shots, and configured to adjust the length of the sub-shots to correspond to the music sub-clips, and to superimpose the syllables of the lyrics of the song over the sub-shots ({SAS} sections of a text track and additional media resources are linked to the time segments, Column 6, Lines 55-57) [Claim 25].

7. Stelovsky teaches an apparatus, comprising: means for creating music sub-clips according to a song, and means for defining and selecting visual content sub-shots (multimedia presentation track consisting of video, audio, and text display is segmented with respect to specific beginning and ending points, Column 3, Lines 27-65); means for timing delivery of syllables of lyrics of the song (sections of the text track are linked to the time segments, Column 6, Line 55; the text can be superimposed on the video, Column 10, Lines 5-6, also Column 14, Lines 52-59 and Column 9, Lines 15-21); and means for assembling the music sub-clips with the visual content sub-shots and adjusting the length of the sub-shots to correspond to the length of the music sub-clips (the music video is synchronized with a song's audio as well as the song's lyrics, and partitioned into time segments that correspond to the song's phrases, Column 8, Lines 34-45), and to superimpose the syllables of the lyrics of the song over the sub-shots (While the song is playing, the corresponding phrases are highlighted in the lyrics field. If necessary, the lyric's field is automatically scrolled to reveal the current phrase, Column 8, Lines 34-45) [Claim 40].

8. What Stelovsky fails to teach is where the segmenting of music to produce a plurality of music sub-clips establishes boundaries between the music sub-clips at beat positions within the music [Claims 1, 23, 25, & 40], and wherein each sub-clip has a duration that is a function of song tempo [Claim 28]. However, Wang teaches a method of detecting beats in a music stream (Beat is defined in the relevant art as a series of perceived pulses dividing a musical signal into intervals of approximately the same duration. Beat detection can be accomplished by any of three methods. The preferred method uses the variance of the music signal, which variance is derived from decoded Inverse Modified Discrete Cosine Transformation (IMDCT) coefficients. The variance method detects primarily strong beats. The second method uses an Envelope scheme to detect both strong beats and offbeats. The third method uses a window-switching

pattern to identify the beats present. The window-switching method detects both strong and weaker beats. In one embodiment, a beat pattern is detected by the variance and the window switching methods. The two results are compared to more conclusively identify the strong beats and the offbeats, Para. 0070-0074; see also Figure 7, the numbered delta functions are understood to be detected beats), and segmenting the music stream at beat boundaries (A normal, error-free audio transmission is represented in the top graph {of Figure 6} by a first and second beat-to-beat interval waveform. The first waveform includes a first beat and the audio information up to a second beat. Similarly, the second waveform includes the second beat and the audio information up to a third beat; In accordance with the method of the present invention, a replacement waveform, including a replacement beat, is copied from the first beat and the first waveform; and is substituted for the missing audio segment in the time interval τ_1 to τ_2, as shown in the bottom graph; all at Para. 0058-0069; see also Figure 6). The beat intervals are taught by Wang to be a function of song tempo (the beat-to-beat interval is replaced by the audio data frames from a corresponding beat-to-beat interval in a preceding 4/4 bar. Most popular music has a rhythm period in 4/4 time, Para. 0067; 4/4 time is understood to be a tempo). Any of the three methods taught by Wang would be used to detect beats in a music clip, and Wang's method of copying and pasting music waveforms segmented by at beat positions would be used to align video, still pictures, music, and lyrics along those boundaries, in the manner as taught by Stelovsky. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have used Wang's methods of segmenting of music to produce a plurality of music sub-clips, establishing boundaries between the music sub-clips at beat positions within the music, with the methods of Stelovsky for integrating lyrics, music, and video content suitable for karaoke, in order to exploit the beat pattern of music

signals to improve the presentation of music when transferred over a network [Claims 1, 23, 25, 28, & 40].

9. What Stelovsky, Wang, and Hansen fail to explicitly teach is where the uniformly distributed sub-shots preserve a storyline represented by the visual content [Claims 1, 23, 25, & 40]. However, Umeda teaches a karaoke authoring apparatus in which the segmented video images may be a series of pictures, scenes, dynamic images, or still pictures presenting a story (Column 4, Lines 23-31). The sub-shots of Stelovsky, selected in a uniform distribution over a timeline of a video, as taught by Hansen, would preserve a chronological story as taught by Umeda. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to preserve a storyline represented by the visual content, as taught by Umeda, in the karaoke system and method of Stelovsky, in light of the teachings of Wang and Hansen, in order to avoid placing sub-shots out of their natural chronological order, such that an order of events is preserved logically [Claims 1, 23, 25, & 40].

10. What Stelovsky, Wang, and Umeda fail to teach is selecting sub-shots such that they are uniformly distributed within the video [Claims 1, 23, 25 & 40]. However, Hansen teaches a system and method for automatically producing media content by creating video subclips called "microchannels" by a “microchannel creator” that determines the desired channel content based upon uniform distribution of video, video and audio, still images and mosaics of different locations (The channel creator then accesses the individual clips from the database and creates the continuous stream or "microchannel." The continuous stream is defined by a concatenated stream of output, whether it be a series of images, video and audio, or other forms of media; The microchannel creator makes the following decisions when creating a microchannel: (i) what type of media should be sent at a given time (video, audio, image); (ii) what triggers should be given priority, assuming multiple triggers are defined for the microchannel; (iii) when advertising

should be inserted into the video stream, and what advertising should be provided; and (iv) when the database should be accessed for pre-recorded clips that are not currently posted to the microchannel as new clips. The channel creator runs via decision algorithms that are determined by the desired channel content for the microchannel. This is best illustrated by example. Considering a hypothetical travel-related site, the following type of microchannel might be desired: (i) commercials should be presented once per minute in ten second maximum durations; (ii) uniform distribution of video, video and audio, still images and mosaics of different locations; (iii) emphasis on video content using activity triggers on beach cams and urban cams; (iv) emphasis on mosaic content using periodic triggering without motion for panoramic cameras; (v) emphasis on still image content for interior cameras, such as restaurant cameras; (vi) live, real-time clips during daylight hours; and (vii) pre-recorded clips during night hours when beach activity has ceased, Para. 0085-88). As best understood, Hansen teaches selecting “microchannels” uniformly from a source. The “microchannel creator” of Hanson would be used in the device of Stelovsky to uniformly select video and photographic content. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to selecting sub-shots such that they are uniformly distributed within the video, as taught by Hansen, in the device of Stelovsky, in light of Wang and Umeda, in order to automatically produce and distribute media content to a targeted audience, for providing more interesting and representative content [Claims 1, 23, 25 & 40].

11. Stelovsky teaches instructions for shortening some of the plurality of sub-shots to a length of a corresponding music sub-clip (the system displays the current segment’s start and end points, so the author can select and edit the boundary points, Column 7, Lines 14-19). Stelovsky teaches instructions for obtaining lyrics from a file (textual track can be generated remotely and transmitted using communications means, Column 14, Lines 20-24); and

coordinating delivery of the lyrics with the music using timing information contained within the file (Column 3, Lines 52-65). What Stelovsky, Wang, Hansen, and Umeda fail to teach is where segmenting is repeated until lengths of all sub-shots are shorter than a maximum of sub-shot length, the maximum of sub-shot length being a little longer in duration than the maximum of music sub-clips [Claim 1]. However, Applicant has not disclosed that having the sub-shots be a "little longer" in duration than the music sub-clips solves any stated problem or is for any particular purpose. Moreover, it appears that the arbitrary length of the sub-clips of Stelovsky or the Applicant's instant invention would perform equally well for synchronizing the sub-clips with a video. Accordingly, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have modified Stelovsky such that lengths of all sub-shots are shorter than a maximum of sub-shot length, the maximum of sub-shot length being a little longer in duration than the maximum of music sub-clips, in light of Wang, Hansen, and Umeda, because such a modification would have been considered a mere design consideration, which fails to patentably distinguish over Stelovsky, Wang, Hansen, or Umeda [Claim 1].

12. What Stelovsky, Wang, Hansen, and Umeda fail to teach is wherein segmenting the visual content comprises instructions for: dividing a shot into two sub-shots at a maximum peak of a frame difference curve; and repeating the dividing to result in sub-shots shorter than a maximum sub-shot length [Claim 1]. However, Golin teaches the use of a Frame Dissimilarity Measure (FDM), which is the ratio of a net dissimilarity measure and a cumulative dissimilarity measure of two consecutive frames (Column 3, Line 65 to Column 4, Line 12). The processing of sub-shots uses the FDM to identify transitions between shots in a video sequence, which appear as peaks in the FDM data (Column 5, Lines 21-42). The data analysis for the sub-shot dividing is a loop, which starts with frames at the beginning of the video sequence and scans through the data to the frames at the end of the sequence (Column 5, Lines 54-62). The length

of the entire video sequence is a maximum sub-shot length. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have used the FDM peak analysis of dividing sub-shots, as described in Golin, for the video segmenting used in Stelovsky, in light of Wang, Hansen, and Umeda, in order to more effectively detect gradual transitions between subshots [Claim 1].

13. What Stelovsky, Wang, Hansen, and Umeda fail to teach is wherein the filtering of a plurality of sub-shots is according to importance or quality [Claim 1]. However, Osberger teaches giving areas of medium motion high importance (Column 7, Lines 10-21). Osberger also teaches that areas of low texture (quality) such as faces are strong attractors of attention (Column 8, Lines 40-54). The sub-shots that are high in “regions of interest”, or attention attracting, are identified (filtered) as taught by Osberger (Column 2, Lines 24-41). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have used the methods of Osburger for filtering sub-shots based on attention indices such as importance to the camera and texture quality, in the karaoke video segmenting device of Stelovsky, Wang, Hansen, and Umeda, in order to increase the entertainment value of the karaoke experience to a user [Claim 1].

14. Stelovsky teaches a karaoke game where the beginning of the track is synchronized with the other tracks of the presentation (9:20-61). The “user’s voice” soundtrack is partitioned into the same time segments as the other tracks. Stelovksy teaches that, using SAS, an author partitions a multimedia presentation into time segments according to predominant time units, e.g., measures of song, sound bites, or action sequences in a movie (6:58-61). Stelovsky teaches where the textual track can be song lyrics (3:41-45). Sections of a text track are linked to each of the time segments (6:62). The multimedia game is recorded onto a mass-storage media (7:3-4). Stelovsky thus teaches where the timing information and the lyric information are

stored in a multimedia game file. What Stelovsky fails to explicitly teach is where the delivery of the lyrics with the music is coordinated with information contained in the lyrics file [Claim 1]. However, Wang clearly teaches the use of MIDI files (1:43-65). For Applicant's benefit, recall that the Musical Instrument Digital Interface (MIDI) format is a protocol that allows computers to control electronic musical instruments. MIDI does not transmit an audio signal or media — it transmits "event messages" such as the pitch and intensity of musical notes, control signals for parameters such as volume, vibrato and panning, cues, and clock signals to set the tempo. MIDI-Karaoke files are an "unofficial" extension of MIDI files, used to add synchronized lyrics to standard MIDI files. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have timing information of Stelovsky would be contained with the textual information in a MIDI file, as taught by Wang, in order for a PC to communicate with a karaoke synthesizer or microphone in the system of Stelovsky [Claim 1].

15. Stelovsky teaches where a sub-shot comprises a video of at least a predetermined length based on the length of a music sub-clip (The recording creates a new "user's voice" sound track. As the beginning of this track is well known, the track is synchronized with the other tracks of the presentation. As a consequence, the "user's voice" sound track is partitioned into the same time segments as the other tracks, Column 9, Lines 31-37). What Stelovsky and Wang further fail to teach is wherein each sub-shot comprises a segment of video of at least a predetermined length based on the length of the music sub-clips and segmented based on a magnitude of difference between adjacent frames [Claim 8]. However, Hansen teaches a system and method for automatically producing media content, in which the clip has a predetermined minimum length {one still frame}, based on detected trigger events in the clip (A "clip" may be defined as a duration of time when the triggers that are set for the capture system are activated--such as when there is motion in the scene and the trigger is set to a basic motion

cue. The clip preferably ends when the trigger event is no longer detected or when a certain time period expires, although other more sophisticated methods for trigger intervals may also be utilized. Once a clip is delineated, the content is generated. At a minimum, the content includes one still image that represents the trigger event in action. For example, 15 seconds out of one minute of captured content may be identified as qualifying content, Para. 0043). What Stelovsky, Wang, Hansen, and Umeda fail to explicitly teach is where the trigger events are based on the length of music sub-clips and segmented based on a magnitude of frame difference [Claim 8]. However, Golin teaches the use of an FDM to segment video (Column 3, Line 65 to Column 4, Line 12; Column 5, Lines 21-42 and Lines 54-62). The FDM of Golin is the magnitude of dissimilarity between two consecutive frames of a video, as demonstrated above. This FDM would be used as a trigger event, as described in Hansen, when used to determine the length of a video sub shot in the system and method of time-segmenting taught by Stelovsky. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have used the frame dissimilarity measure of Golin to determine a sub-shot length in the system and method of Golin, in light of the teachings of Wang, Hansen, and Umeda, in order to synchronize audio tracks with gradual transitions between shots in a video, in order to parse a video for segmentation that does not have abrupt shot transitions [Claim 8].

16. What Stelovsky, Wang, Hansen, and Umeda further fail to explicitly teach is wherein the segmenting music comprises instructions for bounding the sub-clip's length according to: minimum length = min(max(2*tempo,2),4) and maximum length = minimum length+2 [Claim 17], or establishing the music sub-clip's length within a range of 3 to 5 seconds [Claim 18]. However, Applicant has not disclosed that having (min(max(2*tempo,2),4) < length < min(max(2*tempo,2),4)+2) or (3 < length < 5) seconds solves any stated problem or is for any

particular purpose. Moreover, it appears that the arbitrary length of the sub-clips of Stelovsky or the Applicant's instant invention would perform equally well for synchronizing the sub-clips with a video. Accordingly, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have modified Stelovsky such that the music sub-clips had a rigid minimum and maximum length, in light of Wang, Hansen, and Umeda, because such a modification would have been considered a mere design consideration, which fails to patentably distinguish over Stelovsky, Wang, Hansen, or Umeda [Claims 17 & 18].

17. Stelovsky teaches wherein obtaining lyrics comprises instructions for sending the file over a network to a karaoke device (textual track can be generated remotely and transmitted using communications means, Column 14, Lines 20-24; on-line services provide downloading of files, e.g. Internet, Column 6, Lines 49-50) [Claim 24].

18. Stelovsky teaches wherein the visual content analyzer is configured to segment video into sub-shots (Column 6, Lines 51-54) [Claim 29].

19. Stelovsky teaches wherein the means for defining and selecting visual content sub-shots is a video analyzer configured to segment video into sub-shots (Using SAS, the author partitions the multimedia presentation into time segments according to predominant time units, e.g., action sequences in a movie, Column 6, Lines 51-54) [Claim 41].

20. What Stelovsky, Wang, Hansen, and Umeda further fail to teach is wherein filtering the plurality of sub-shots according to importance comprises instructions for evaluating frames within a sub-shot according to attention indices, and averaging the attention indices for the frames to determine if the sub-shot should be included [Claim 6]. However, Osberger teaches identifying and adaptively segmenting frames of video based upon an attention model, AKA total importance map, composed by linear weighting of the spatial and temporal importance maps (Column 2, Lines 24-41). It is inherent that averaging is merely linear weighting with a weight

factor of one. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have utilized the averaging of the attention indices of Osberger to select frames of importance, for use in the karaoke system of Stelovsky, in light of Wang Hansen, and Umeda, in order to adapt the attention model for a variety of different types of video sub-shots, while accurately determining regions of interest in the videos [Claim 6].

21. What Stelovsky, Wang, Hansen, and Umeda further fail to teach is wherein filtering the sub-shots according to importance comprises instructions for analyzing the camera motion, object motion, and specific objects within the subshots, and filtering the subshots according to the analysis [Claim 7], or wherein a visual content analyzer is configured to select from the sub-shots according to ranked importance, gauged by detection of color entropy, object motion, camera motion, or of a face within the sub-shot [Claims 10 & 32]. However, Osberger teaches selecting or filtering sub-shots by color information (Column 3, Lines 6-15), by camera or object motion (Column 7, Lines 7-37), or by specific objects, including faces, in a sub-shot (Column 8, Lines 40-54). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have used the various color, motion, and object detection in the video sub-shots, as described by Osberger, in the personalized karaoke system on Stelovsky, in light of Wang Hansen, and Umeda, in order to improve the prediction of visual importance of a sub-shot [Claims 7, 10, & 32].

22. What Stelovsky, Wang, Hansen, and Umeda further fail to teach is wherein filtering the plurality of sub-shots comprises instructions for: examining color entropy within each of the plurality of sub-shots to detect motion more than a threshold indicating interest and less than a threshold indicating low camera and/or object movement; and selecting sub-shots having acceptable motion and/or color entropy scores [Claim 5], or wherein the visual content analyzer is configured to filter out sub-shots having low image quality, as measured by low entropy and

low motion intensity [Claim 33]. However, Osberger teaches segmenting frames into regions based upon both color and luminance (Column 2, Lines 24-41). The term entropy is taken to mean Information Entropy or Shannon Entropy, which refers to a measure of uncertainty associated with a random variable. Thus, referring to lossless data compression, the color entropy would refer to an average minimum number of bits needed to communicate a color value. Osberger teaches using an algorithm to segment an image into homogeneous regions using color information, to generate the spatial importance map (Column 3, Lines 6-15). Osberger also teaches that, if the spatial importance map is too noisy from frame to frame, a temporal smoothing operation is performed, and a temporal importance map is generated (Column 6, Line 66 to Column 7, Line 37). The temporal importance map is calculated using adaptable thresholds because the amount of motion varies greatly across different scenes. Osberger also teaches identifying sub-shots with regions of interest by using the spatial and temporal interest maps in order to produce an adaptive segmentation model (Column 8, Lines 58-67), for segmenting video scenes. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have incorporated the color entropy detection, then the camera motion detection of Osberger with the segmentation of karaoke video as described by Stelovsky, in light of Wang, Hansen, and Umeda, in order to attract the interest of a karaoke user more effectively [Claims 5 & 33].

23. Claims 12-15, 31, 34-38, & 43 are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Wang, Hansen, Umeda, Golin, and Osberger, and further in view of Geigel et al. (US 2002/0122067 A1), hereinafter known as Geigel.

24. Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger teach all the features as demonstrated in the rejection of claims 1, 25, & 40 above. What Stelovsky, Wang, Hansen,

Umeda, Golin, and Osberger fail to explicitly teach is wherein the instructions for segmenting visual content includes assigning photographs to be sub-shots [Claim 12], instructions for assigning photographs includes converting at least one photograph to video [Claim 14], wherein the visual content comprises home video and photographs in digital formats [Claim 15], wherein a visual content analyzer is configured to assemble still photographs, each of which is a sub-shot [Claim 31], and wherein the visual content analyzer is configured to define sub-shots from visual content comprising photographic and video content [Claim 34]. However, Geigel teaches a layout generator for digital images (Para. 0010), including photographs or video clips (Para. 0055), which converts the images into a video (output is Picture CD media or other photo delivery media, Para. 0057). It is inherent that a series of images displayed during a progression of time is a video. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have assembled and converted photos to video, as taught by Geigel, for the background video in the entertainment system of Stelovsky, in light of Wang, Hansen, Umeda, Golin, and Osberger, in order to automate the layout of the background in a manner pleasing to the user [Claims 12, 14, 15, 31, & 34].

25. What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger further fail to teach is wherein a visual content analyzer is configured with instructions for assigning photographs includes instructions for: rejecting photographs having problems with quality [Claim 13]; and rejecting a similar group of photographs when one within the group has been selected [Claims 13 & 37]. However, Geigel teaches performing detection of dud images and duplicate images prior to being submitted to the layout system (Para. 0061). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have not selected dud or duplicate images when creating the background image layout, as shown by Geigel, when implementing the entertainment system of Stelovsky, in light of Wang, Hansen,

Umeda, Golin, and Osberger, in order to necessitate the minimal input from the user when assembling images aesthetically pleasing to the user [Claims 13 & 37].

26. What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger further fail to teach is wherein a visual content analyzer is configured to organize photographs by the date of exposure and scene, thereby obtaining photographs having a relationship [Claim 36]. However, Geigel teaches organizing the images (page layout algorithm, Para 0059) by date of exposure (chronology of the images, Para. 0063) and scene (event clustering, Para. 0060). It is inherent that all the photographs would thus be related by a date range or event group. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have organized the images to the extent provided by Geigel, is the operation of the entertainment system of Stelovsky, in light of Wang, Hansen, Umeda, Golin, and Osberger, in order to distribute the photographs automatically according to an algorithm that valued a user-pleasing arrangement [Claim 36].

27. What Stelovsky, Wang, Hansen, and Umeda fail to teach is wherein a visual content analyzer is configured to reject photographs of low quality by detecting over and under exposure, overly homogeneous images, and blurred images [Claim 35]. Osberger teaches a visual analyzer (image processing algorithm) to detect overexposure and underexposure (contrast), overly homogeneous images (homogeneous regions, Column 3, Lines 6-15), and blurred images (areas of very high motion, Column 7, Lines 10-26). What Stelovsky, Wang, Hansen, and Osberger fail to teach is wherein the visual content analyzer rejects photographs which are underexposed, overexposed, overly homogeneous, or blurred [Claim 35]. However, Geigel teaches selection of the best image (Para. 0057). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have rejected images which are underexposed, overexposed, overly homogeneous, or blurred, in light of the

teachings of Osberger and Geigel, in the entertainment system of Stelovsky, in light of Wang and Hansen, in order to discriminate images to present highly desirable visuals to a karaoke user [Claim 35].

28. What Stelovsky, Wang, Hansen, and Umeda further fail to teach is wherein the means for defining and selecting visual content sub-shots is a video analyzer configured for: detecting an attention area within a photograph; and creating a photo to video sub-shot based on the attention area, wherein the video includes panning and zooming [Claims 38 & 43]. Osberger teaches a visual analyzer (image processing algorithm) to detect an attention area within a photograph (Column 2, Lines 24-41), and wherein motion vectors are used by camera motion estimation algorithm to determine pan and zoom in a frame (Column 7, Lines 22-37). What Stelovsky, Wang, Hansen, and Osberger fail to teach is wherein photo to video subshot includes panning and zooming. However, Geigel teaches, in photography terms rather than videography terms, panning the images (auto-cropping, Para. 0057) and zooming the images (scaling, Para. 0122). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to created a photo to video sub-shot based on a detected attention area, including panning and zooming, in light of the teachings of Osberger and Geigel, in the entertainment system of Stelovsky, in light of Wang, Hansen, and Umeda, in order to further refine the content information of an image by focusing on the attention-attracting elements in the photo to video, when used as the background for karaoke entertainment [Claims 38 & 43].

29. Claims 19, 39, & 44 are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Wang, Hansen, Umeda, Golin, and Osberger, and further in view of Bloom et al. (US 2005/0042591 A1), hereinafter known as Bloom.

30. Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger teach all the features as demonstrated above in the rejections of claims 1, 18, 25, & 40 above, including wherein the lyric formatter is configured to consume a file detailing timing of the lyrics (the textual track can be generated remotely and transmitted by communication means, digitally, using a software program, Column 14, Lines 14-24; the digital textual track used for the karaoke is inherently a file to be "consumed" or used). Stelovsky teaches wherein evaluation of output can involve differences in pronunciation patterns and any processes involved in generating speech (Column 14, Lines 52-59). What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger fail to teach is wherein segmenting the music comprises a lyric formatter configured with instructions for establishing boundaries for the music sub-clips at sentence breaks [Claim 19], or consuming a file detailing timing of each syllable and each sentence of the lyrics [Claims 39 & 44], and for rendering the lyrics syllable by syllable [Claim 44]. However, Bloom teaches automatically synchronizing sound to images, wherein lyric segmentation may be syllable by syllable (line can be a single word or sound) or a sentence (Para. 0139). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have segmented the music of the karaoke system of Stelovsky, in light of the syllable and sentence boundaries of the lyrics as taught by Bloom, in light of Wang, Hansen, Umeda, Golin, and Osberger, in order to synchronize the song with a user's lip movements on the accompanying video display [Claims 19, 39, & 44].

31. Claim 21 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Wang, Hansen, and Umeda, and further in view of Tsai (US 6,572,381 B1), hereinafter known as Tsai.

32. Stelovsky, Wang, Hansen, and Umeda teach all the features as demonstrated above in the rejections of claims 1 & 20 above. What Stelovsky, Wang, Hansen, and Umeda fail to teach is wherein obtaining the lyrics comprises instructions for sending the file over a network to a karaoke device as part of a pay-for-play service [Claim 21]. However, Tsai teaches a plurality of karaoke terminals connected to a host computer via a network (communications line) that delivers lyric data (Column 8, Lines 48-61). Tsai teaches a karaoke system shares the source data as part of a pay service (Column 2, Lines 48-56; also Column 20, Line 52 to Column 21, Line 56). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have sent the lyrics file over a network in conjunction with a pay-for-play service, as taught by Tsai, in the karaoke system of Stelovsky, in light of Wang Hansen, and Umeda, in order to offer commercial messages with updated custom content to a subscriber of a karaoke service [Claim 21].

33. Claim 22 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Wang, Hansen, Umeda, Golin, and Osberger, and further in view of Tashiro et al. (US 5,703,308), hereinafter known as Tashiro.

34. Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger teach all the features as demonstrated above in the rejections of claim 1 above. What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger fail to teach is wherein the processor-readable medium comprises instructions for: querying a database of songs by humming a portion of a desired song; and selecting the desired song from among a number of possibilities suggested by an interface to the database [Claim 22]. However, Tashiro teaches a karaoke device having database of songs (music data storage device with a plurality of entry songs stored in a data table, Column 1, Line 54 to Column 2, Line 3), wherein the database is queried by humming a song (key melody

patterns which represent a desired song are input by voice, Column 3, Lines 10-14) and selecting the desired song through an interface (music selection is made from top 10 matching entries, Column 7, Lines 48-67). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, in the karaoke system of Stelovsky, to search and select a desired song from a database by humming, as taught by Tashiro, in light of Wang Hansen, and Umeda, in order to select a song even if neither the artist nor the title of the song is known [Claim 22].

35. Claim 26 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Wang, Hansen, Umeda, Golin, and Osberger, and further in view of Trovato et al. (US 7,058,889 B2), hereinafter known as Trovato.

36. Stelovsky, Wang, Hansen, and Umeda teach all the features as demonstrated above in the rejections of claims 1 & 25. What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger fail to teach wherein the music analyzer is configured to segment the song with a strong onset between each of the music sub-clips [Claim 26]. However, Trovato teaches locating transition points for a music segmentation scheme by onset break detection (Column 7, Lines 33-51; also Figure 6). It is inherent from Figure 6 that weak onset breaks are not used as transition points. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have analyzed the music used in the karaoke system of Stelovsky with the onset break detection method defined in Trovato, in light of Wang, Hansen, Umeda, Golin, and Osberger, in order to automatically synchronize the music with the background video consistent with human perception [Claim 26].

37. Claim 27 is rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, in view of Wang, Hansen, Umeda, Golin, and Osberger, and further in view of Kondo (US 6,232,540 B1), hereinafter known as Kondo.

38. Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger teach all the features as demonstrated above in the rejections of claims 1 & 25. What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger fail to teach is wherein a music analyzer is configured to segment the music automatically, comprising instructions for: establishing boundaries for the music sub-clips with a beat position between each of the music sub-clips [Claim 27]. However, Kondo teaches establishing boundaries (positions) for music sub-clips (rhythm sound source signals) at beat positions within the music (positions of attacks in the rhythm sounds, Abstract). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have divided the music sub-clips at beat positions within the music, as shown in Kondo, for use in the karaoke system of Stelovsky, in light of Wang Hansen, and Umeda, in order to avoid occurrences of rhythm disorder in the rhythm sounds [Claim 27].

39. Claims 30 & 42 are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger, in view of Borden, IV et al. (US 2003/0200105 A1), hereinafter known as Borden IV.

40. Stelovsky, Wang, Hansen, and Umeda teach all the features of claims 25 & 40 above. What Stelovsky, Wang, Hansen, and Umeda fail to teach is where the video analyzer or visual content analyzer is configured to access folders of home video and photographs containing content from which the sub-shots are derived [Claims 30 & 42]. However, Border IV teaches a video analyzer (user's data processing device) which can access folders of a customer's video or photographs (MY PHOTOS homepage document, containing a user's uploaded images or

video, Para. 0016-0017). Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have accessed a user's personal video and photo content for generating the sub-shots, in the karaoke device of Stelovsky, in light of Wang Hansen, and Umeda, in order to attract potential customers to receive services by hosting their personal data [Claims 30 & 42].

41. Claims 11 & 45 are rejected under 35 U.S.C. 103(a) as being unpatentable over Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger, as applied to claim 1 above, and further in view of Haitsma et al. (US 2002/0178410 A1), hereinafter known as Haitsma.

42. Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger teach all the features of claim 1 as demonstrated above. What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger fail to teach is wherein the selecting uniformly distributed sub-shots comprises evaluating a normalized entropy of the sub-shots along a time line of video from which the sub-shots are obtained [Claim 11]. However, Haitsma teaches a hashing method for indexing video clips in a database, in which a normal distribution is calculated for video clips to determine whether they are different quality versions of the same content (Two 3 seconds audio clips (or two 30-frame video sequences) are declared similar if the Hamming distance between the two derived hash blocks H.sub.1 and H.sub.2 is below a certain threshold T. This threshold T directly determines the false positive rate P.sub.f, i.e. the rate at which two audio clips/video sequences are incorrectly declared equal (i.e. incorrectly in the eyes of a human beholder): the smaller T, the smaller the probability P.sub.f will be. On the other hand, a small value T will negatively effect the false negative probability P.sub.n, i.e. the probability that two signals are `equal`, but not identified as such. In order to analyze the choice of this threshold T, we assume that the hash extraction process yields random i.i.d. (independent and identically distributed) bits. The number

of bit errors will then have a binomial distribution with parameters (n,p), where n equals the number of bits extracted and p(=0.5) is the probability that a `0` or `1` bit is extracted. Since n(32.times.256=8192 for audio, 32.times.30=960 for video) is large in our application, the binomial distribution can be approximated by a normal distribution with a mean .mu.=np and standard deviation .sigma.=[square root][square root over (np(1-p))], Para. 0041). This is understood to be a normalized entropy in the sense that the normal video quality is used to determine the similarity of sub-shots. Such a method would be used in the system and method of Stelovsky to determine whether a video clip or photograph duplicates the content of another except in quality. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have selected a uniform distribution of sub-shots along a timeline, as taught by Hansen, by analyzing the normalized entropy of the sub-shots, as taught by Haitsma, in light of the teachings of Wang, Hansen, Umeda, Golin, and Osberger, in order to avoid the non-uniform selection of duplicate sub-shot content in sub-shots that have distinct data representations due to differing quality [Claim 11].

43. Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger teach all the features of claim 40 as demonstrated above. Stelovsky teaches means for displaying assembled visual content comprising sub-shots with music sub-clips (Column 3, Lines 27-41). Hansen teaches wherein the means for defining and selecting visual content sub-shots is such that the sub-shots are uniformly distributed within the visual content (Para. 0085-88). What Stelovsky, Wang, Hansen, Umeda, Golin, and Osberger fail to teach is where the sub-shots are uniformly distributed within the visual content is further configured for selecting uniformly distributed sub-shots via evaluating normalized entropy of the sub-shots along a time line of visual content from which the sub-shots were obtained [Claim 45]. However, Haitsma teaches a hashing method for indexing video clips in a database, in which a normal distribution is calculated for video clips to

determine whether they are different quality versions of the same content (Para. 0041). This is understood to be normalized entropy in the sense that the normal video quality is used to determine the similarity of sub-shots. Such a method would be used in the system and method of Stelovsky to determine whether a video clip or photograph duplicates the content of another except in quality. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to have selected a uniform distribution of sub-shots along a timeline, as taught by Hansen, by analyzing the normalized entropy of the sub-shots, as taught by Haitsma, in light of the teachings of Wang, Hansen, Umeda, Golin, and Osberger, in order to avoid the non-uniform selection of duplicate sub-shot content in sub-shots that have distinct data representations due to differing quality [Claim 45].

44. What Stelovsky, Wang, Hansen, Golin, and Osberger further fail to teach is where the means for displaying the assembled visual content comprising sub-shots with music sub-clips is configured such that displaying the assembled visual content preserves a storyline as represented by the visual content [Claim 45]. However, Umeda teaches a karaoke authoring apparatus in which the segmented video images may be a series of pictures, scenes, dynamic images, or still pictures presenting a story (Column 4, Lines 23-31). The sub-shots of Stelovsky, selected in a uniform distribution over a timeline of a video, as taught by Hansen, would preserve a chronological story as taught by Umeda. Therefore, it would have been obvious to one of ordinary skill in the art, at the time the invention was made, to preserve a storyline represented by the visual content, as taught by Umeda, in the karaoke system and method of Stelovsky, in light of the teachings of Wang, Hansen, Golin, and Osberger, and Haitsma, in order to avoid placing sub-shots out of their natural chronological order, such that an order of events is preserved logically [Claim 45].

Response to Arguments

45. Applicant's arguments filed 2/5/2009 have been considered and are not persuasive for the following reasons. Applicant argues at page 27, Para. 0023-24 of the response that Examiner equates 4/4 time [signature] with tempo. To clarify the rejection, the limitation recited in Claims 1, 23, 25, & 40 is "wherein the beat positions are located according to a rhythm or a tempo of the music." The additional verbiage of only Claim 1 recites "…or at onset positions within the music when beat positions are not obvious during a portion of the music, the onset positions being initiations of distinguishable tones of the portion of the music…" The language is interpreted by Examiner as in the *alternative*; further that the beat positions are merely located *according to* either a rhythm, a tempo, or onset positions within the music, rather than *at* them or *on* them. Applicant discloses that video shots and photographs are aligned with boundaries defined by the musical beat –i.e., make the video transition happen at the beat positions of the incidental music, whose boundary is at the beat position. See Applicant's specification at page 5, Para. 0025, and page 16, Para. 0053. Stelovsky teaches a Segmentation Authoring System partitions a multimedia track with respect to specific beginning and end points (3:27-65). Stelovsky further suggests that tracks of any media format, such as motion video, audio, sequence of still images, or text can be associated with a multimedia presentation, and be synchronized with respect to the presentation's time or its segments or be independent of its time axis, and that such tracks can be recorded by the user (3:64-4:3). What is not taught by Stelovsky is where these beginning and end points are located according to either a rhythm, a tempo, or an onset position of music. However, Wang teaches a method for detecting beats in a music stream, using a variance method, an envelope scheme, and a window-switching method to detect strong, weak, and offset beats (Para. 0070-74 and Figure 7). The waveform analyzer of Wang would be the preferred method of segmenting music for the method of Stelovsky,

where a user's recorded voice input (i.e., singing) would need to be synchronized *on the beat* of a musical track, in order to allow the a user to create a new vocal track in real time by singing into the microphone (as suggested by Stelovsky at 1:17-21). Examiner understands that a tempo, such as "120 beats per minute" or "allegro" is distinct from a time signature, such as "4/4 time"; however, it is well known that time signatures describe the number of beats in a bar or measure and the note value which represents one beat. For example, "4/4 time" best describes a piece having quadruple quarter beats (en.wikipedia.org/wiki/Time_signature). As best understood, Wang detects beats wherever they appear in a music track; on measured beat positions, and hence according to all of rhythm, tempo, and onset positions. Wang describes detecting strong beats and offbeats; these offbeats are understood to be "unobvious" beat positions. Wang's method of detecting the variance of the music is analogous to distinguishing an initiation of tones in the music. Also, any routineer to karaoke would realize that the lyrics of a western song in common 4/4 time would be sung in coordination with beat positions in the music; and that in a music video, scenes transitions are made on beat positions. For these reasons, Examiner's position is that Wang's beat detector would be used to detect segmentation points for Stelovsky's music analyzer, in a karaoke device, in order to combine the entertainment value of music videos with the functional value of computerized games, for demonstrating how to sing a song; hence, Applicant's argument is unpersuasive because the claim is worded broader than argued and the prior art teaches both method and reasonable motivation for performing steps as disclosed in the specification.

46. Applicant further argues at page 27, Para. 0025-26 that Umeda fails to present a story line or where a story line might be preserved. However, it is common knowledge that movies, music videos, musical theater, etc. may have a story line. Umeda teaches a *karaoke authoring apparatus* in which segmented videos may be a series of scenes or images presenting a story

line (4:29-31). It is obvious that authoring a karaoke music video presents a story line, and that segmenting or editing as such shall preserve the storyline for the enjoyment of the viewer; and this would be done by merely preserving the chronological order to the images. It would be elementary to use this method with the Segmentation Authoring System taught by Stelovsky. Therefore, Applicant’s argument is unpersuasive.

47. Applicant further argues at pages 27-28, Para. 0027-29 that none of the references teach coordinating the delivery of the lyrics with the music using timing information contained within the file. However, Stelovsky at 9:20-61 teaches a karaoke game where the beginning of the track is synchronized with the other tracks of the presentation. The “user’s voice” soundtrack is partitioned into the same time segments as the other tracks. Stelovksy teaches that, using SAS, an author partitions a multimedia presentation into time segments according to predominant time units, e.g., measures of song, sound bites, or action sequences in a movie (6:58-61). Stelovsky teaches where the textual track can be song lyrics (3:41-45). Sections of a text track are linked to each of the time segments (6:62). The multimedia game is recorded onto a mass-storage media (7:3-4). Stelovsky thus teaches where the timing information and the lyric information are stored in a multimedia game file. Applicant’s argument that the presentation time is not stored in a text file is further not persuasive, because Wang clearly teaches the use of MIDI files (1:43-65). For Applicant’s benefit, recall that the Musical Instrument Digital Interface (MIDI) format is a protocol that allows computers to control electronic musical instruments. MIDI does not transmit an audio signal or media — it transmits "event messages" such as the pitch and intensity of musical notes, control signals for parameters such as volume, vibrato and panning, cues, and clock signals to set the tempo. MIDI-Karaoke files are an "unofficial" extension of MIDI files, used to add synchronized lyrics to standard MIDI files (en.wikipedia.org/wiki/Musical_Instrument_Digital_Interface). Thus, it is well established and

inherent in the prior art that timing information would be contained with the textual information in a MIDI file, in order for a PC to communicate with a karaoke synthesizer or microphone in the system of Stelovsky. As such, Applicant's arguments are not persuasive.

48. Applicant argues at page 28, Para. 0031 that none of the cited references discloses “the maximum of sub-shot length being a little longer in duration that the maximum of music sub-clips" and “shortening some of the plurality of sub-shots to a length of a corresponding music sub-clip from within the plurality of music sub-clips.” Applicant further argues at pages 31-32, Para. 0041-46, that neither Stelovsky, Wang, nor Hansen teaches bounding the music sub-clip minimum length according to, as best understood by Examiner, the minimum of the maximum of twice the tempo and 2 seconds, and 4 seconds; and bounding the maximum length to the minimum length plus two seconds; also that establishing the music sub-clips' length to within 3 to 5 seconds is for the particular purpose of “giving a more enjoyable karaoke performance.” However, Stelovsky teaches a system for segmenting music in order to combine the entertainment value of music videos with the informational and educational value of computerized games. Applicant's specification at page 17, Para. 0055-56 do not provide motivation for why the specific length in seconds of the music sub-clip needs to be 3 to 5 in order for the user to enjoy the presentation; as if music sub-clips between 4-6 seconds are terrible. Applicant's specification at pages 27-28, Para. 0091 further fails to explain how the specific formula for length in seconds of the music sub-clips causes the karaoke performance of Applicant's instant invention to be "more enjoyable" than Stelovsky's. Applicant's specification at page 28, paragraph 0092 suggests that the music sub-clip length may be set to within a fixed range, such as 3 to 5 seconds, or may be fine-tuned as desired; indicating that Applicant considers the length to be arbitrary. Examiner further notes that it is mere design choice all round to experiment with segmenting the music sub-clips at various lengths in order to

determine what is "more enjoyable" to a user, as this precise experimentation is what is taught by Stelovsky's authoring system. Thus, it would be obvious to one of ordinary skill in the art of presentation editing to merely try fitting different short lengths of music to lengths of segmented video, in order to please a user. As such, all the limitations are construed to be mere design choice because a user would enjoy the music sub-clips of the length he set in Stelovsky's system. Thus, Applicant's argument is not convincing.

Conclusion

49. Applicant's amendment necessitated the new ground(s) of rejection presented in this Office action. Accordingly, **THIS ACTION IS MADE FINAL**. See MPEP § 706.07(a). Applicant is reminded of the extension of time policy as set forth in 37 CFR 1.136(a).

A shortened statutory period for reply to this final action is set to expire THREE MONTHS from the mailing date of this action. In the event a first reply is filed within TWO MONTHS of the mailing date of this final action and the advisory action is not mailed until after the end of the THREE-MONTH shortened statutory period, then the shortened statutory period will expire on the date the advisory action is mailed, and any extension fee pursuant to 37 CFR 1.136(a) will be calculated from the mailing date of the advisory action. In no event, however, will the statutory period for reply expire later than SIX MONTHS from the date of this final action.

Any inquiry concerning this communication or earlier communications from the examiner should be directed to NIKOLAI A. GISHNOCK whose telephone number is (571)272-1420. The examiner can normally be reached on M-F 11:00a-7:30p EST (8:00a-4:30p PST).

If attempts to reach the examiner by telephone are unsuccessful, the examiner's supervisor, Xuan M. Thai can be reached on 571-272-7147. The fax phone number for the organization where this application or proceeding is assigned is 571-273-8300.

Information regarding the status of an application may be obtained from the Patent Application Information Retrieval (PAIR) system. Status information for published applications may be obtained from either Private PAIR or Public PAIR. Status information for unpublished applications is available through Private PAIR only. For more information about the PAIR system, see http://pair-direct.uspto.gov. Should you have questions on access to the Private PAIR system, contact the Electronic Business Center (EBC) at 866-217-9197 (toll-free). If you would like assistance from a USPTO Customer Service Representative or access to the automated information system, call 800-786-9199 (IN USA OR CANADA) or 571-272-1000.

5/10/2009
/N. A. G./
Examiner, Art Unit 3715

/XUAN M. THAI/
Supervisory Patent Examiner, Art Unit 3715